

Supporting Information

Identification of RNA Pseudoknot-Binding Ligand That Inhibits the -1 Ribosomal Frameshifting of SARS-Coronavirus by Structure-based Virtual Screening

So-Jung Park [†], Yang-Gyun Kim[‡] and Hyun-Ju Park ^{*†}

[†] *School of Pharmacy and* [‡]*Department of Chemistry, Sungkyunkwan University, Suwon 440-746, Korea*

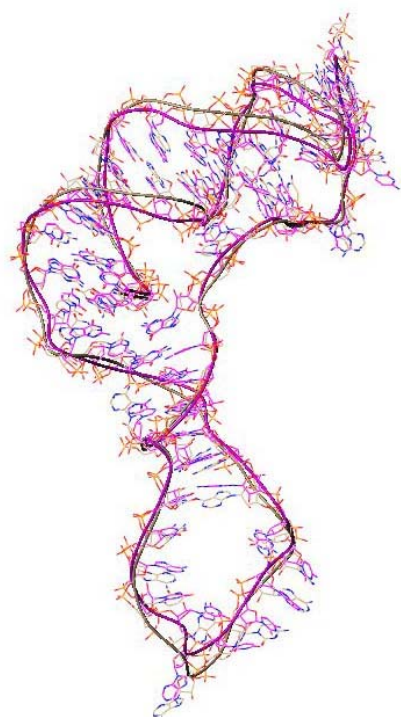
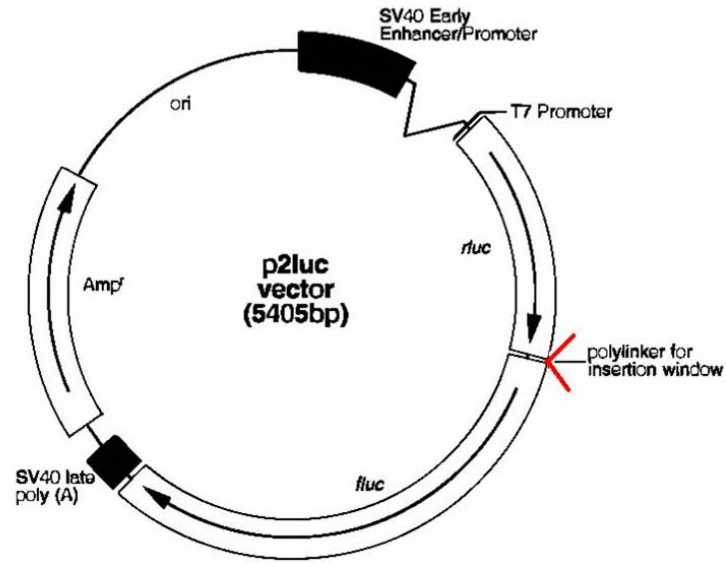


Figure S1. Comparison between averaged structure of SARS-pseudoknot obtained from molecular dynamics simulation after being neutralized by Na^+ (brown) and Mg^{2+} (magenta). Brown tube is averaged structure obtained from MD result after being neutralized by Na^+ . Magenta tube is average structure obtained from MD after being neutralized by 33 Mg^{2+} counterions under the same condition of minimization and molecular dynamics described as in the experimental section. To test the difference of averaged two structures neutralized by Na^+ and Mg^{2+} , rmsd was measured between these two structures (rmsd = 1.65Å). This result suggested that Mg^{2+} simulation do not give significant change in overall structure. (We also docked **43** to averaged structure obtained from Mg^{2+} simulation. **43** docked in to the active site with similar orientation of docking model using Na^+ simulation structure. The nitrogen atom in the thiazole ring in **43** forms a hydrogen bond (2.7Å) with the 2'-OH group of ribose of C62.)

(a)



(b)

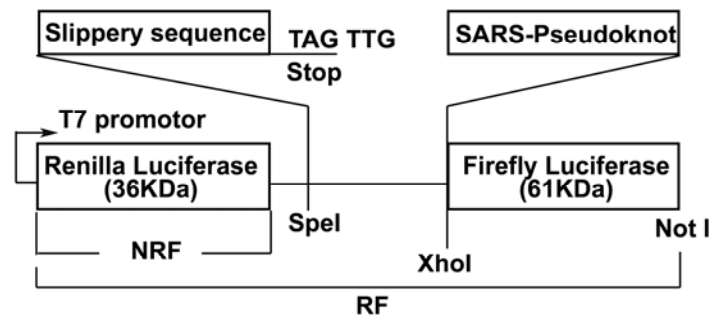


Figure S2. Template construct for -1 RF assay. (a) Dual report vector p2luc system was used in the study for testing -1 RF efficiencies. To make constructs for testing -1 frameshifting (RF), (b) was inserted between XhoI and SpeI restriction sites of the p2luc¹. Alternatively, DNA templates containing a T7 promoter were transcribed and translated using the TNT T7-transcription/translation coupled reticulocyte lysate. Rluc; Renilla luciferase gene, fluc; firefly luciferase gene, Amp^r; ampicillin resistance, ori; origin of replication in Escherichia coli+. (b) A stop codon is found immediately after the slippery site. If a -1 RF occurs at the slippery sequence, the termination codon of the renilla luciferase gene is not read and rluc-fluc fusion protein (RF) was produced. Whereas -1 RF does not occur the termination codon TAG of the renilla luciferase gene (rluc) is read and only renilla luciferase (NRF) was produced.

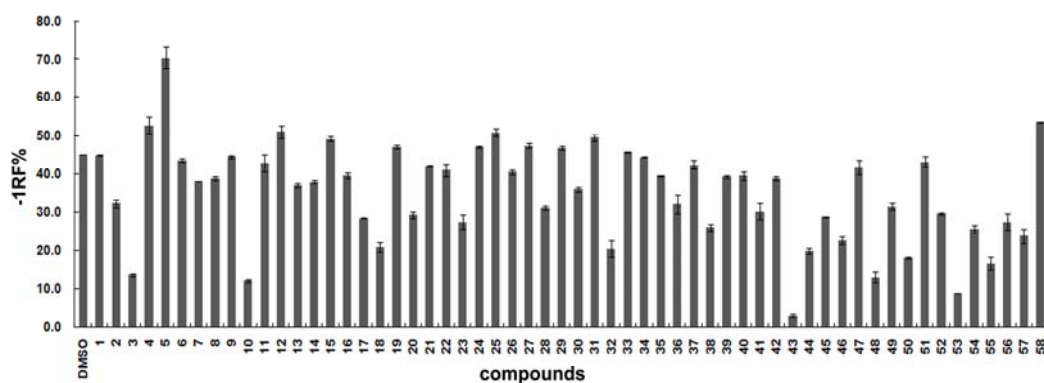


Figure S3. A histogram of -1 RF efficiencies of all compounds obtained by SDS-PAGE analysis after TNT assay. SDS-PAGE analysis of [^{35}S] methionine-labeled translation products from -1 RF assay of p2Luc-SARS pseudoknot in reticulocyte lysate. Each candidate compound was treated at a concentration of 250 μM . The reaction with only DMSO (lane1) considered as a control.

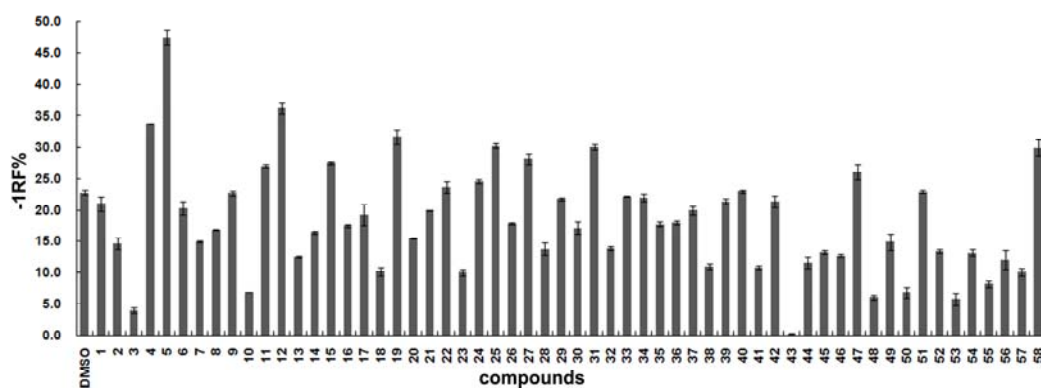


Figure S4. A histogram of the results obtained by dual luciferase assays. Each candidate compounds was added to a concentration of 250 μ M. The reaction with only DMSO (lane1) considered as a control. The activities of firefly and renilla luciferases are measured sequentially from a single sample. Plasmid p2luc that expresses in-frame renilla-firefly fusion protein was used as a positive control for -1 RF (defined as 100% efficiency). -1 RF efficiency was calculated with the formula $\% = [(\text{firefly luciferase of sample} / \text{renilla luciferase of sample}) / (\text{firefly luciferase of p2luci} / \text{renilla luciferase of p2luci})] \times 100$.

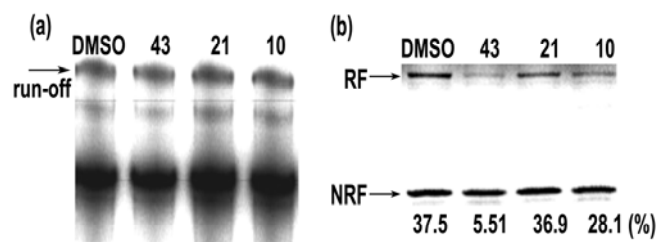
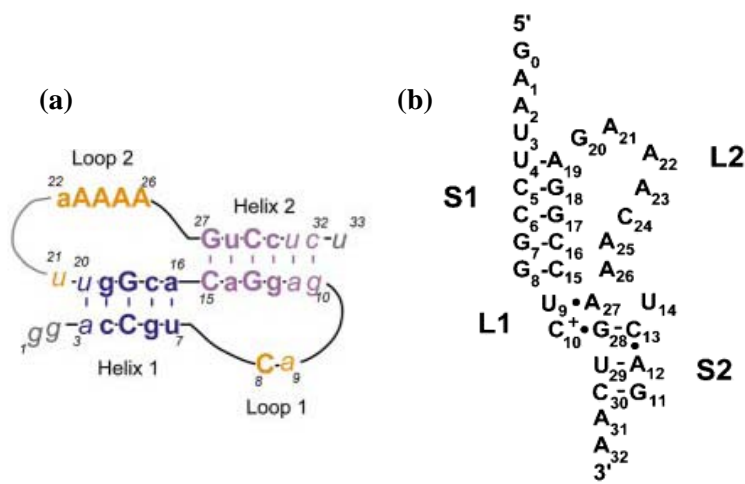


Figure S5. (a) *In vitro* transcription in the presence of compounds, **43**, **21** and **10**. Transcripts synthesized with [α - p^{32}]CTP were analyzed by a 5%-denaturing polyacrylamide gel electrophoresis. Each compound was added at a concentration of 250 μ M. Reaction only with DMSO (lane1) considered as a control. (b) Measurement of -1 RF efficiency using *in vitro* translation assay. Proteins analyzed by SDS-PAGE in the same way as *in vitro* TNT assay. The nonframeshifting product (NRF) is the *Renilla* luciferase protein, and the frameshifting product (RF) is a firefly luciferase-*renilla* luciferase fusion protein.



(c)

p2luc-biotin-pseudoknot; GGACCGUCAGAGGACACGGUUA AAAAAGUCCUCU
p2luc-PEMV; UCCGGUCGACUCCGGAGAAACAAAGUC

Figure S6. Structures and sequences of biotin-pseudoknot and PEMV-pseudoknot. (a) X-ray crystal structure of biotin-binding RNA pseudoknot.¹ (b) The NMR structure of PEMV-RNA pseudoknot.² (c) Sequences of RNA pseudoknots inserted into the p2luc construct for -1 RF assay.

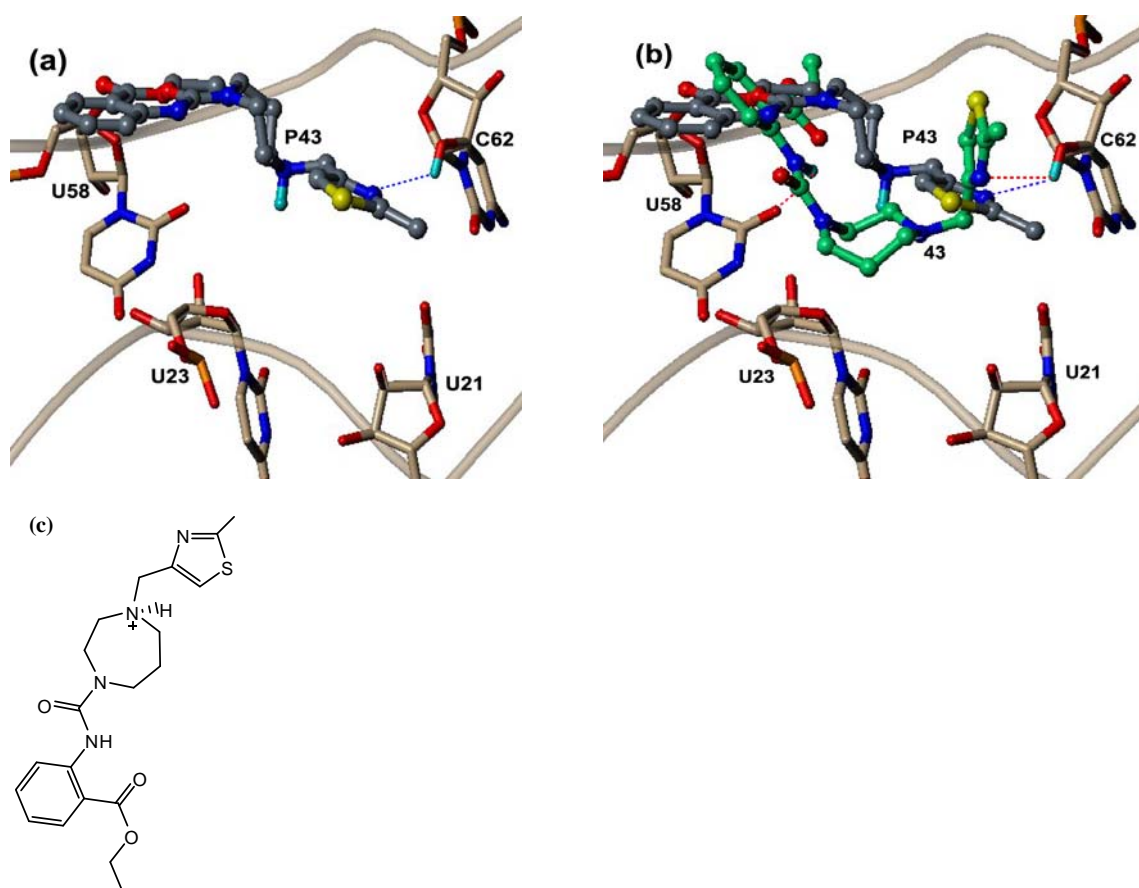


Figure S7. (a) Docked model of protonated **43** (**P43**):SARS-psudoknot complex generated by DOCK 4.0. Several residues in the binding site are rendered in capped stick (brown carbon), protonated **43** (**P43**) in ball and stick (grey carbon) and blue dashed line indicates hydrogen bonds. The nitrogen atom in the thiazole ring in **P43** forms a hydrogen bond (1.8Å) with the 2'-OH group of ribose of C62. (b) Superposition of **43** and **P43** (protonated **43**) docking models obtained from DOCK4.0. Compound **43** is rendered in ball and stick (green carbon) and red dashed lines indicate hydrogen bonds. (c) Structure of protonated **43**.

Certificate of Analysis

Chemical Name : 2-{{[4-(2-Methyl-thiazol-4-ylmethyl)-
[1,4]diazepane-1-carbonyl]-amino}-
benzoic acid ethyl ester

Chemical Formula : C₂₀H₂₆N₄O₃S

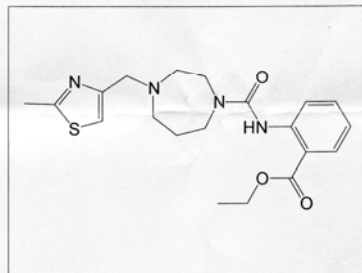
Molecular Weight : 402.51

CAS Number :

Lot. Number : LDY04-179

Synthesized Date : 2005. 12. 24

Quantity : 1.135 G



<u>ANALYSIS ITEM</u>	<u>SPECIFICATION</u>	<u>RESULTS</u>
Appearance	Ivory Oil	Ivory Oil
Assay (by HPLC)	>95% Min.	> 98 %
Identification	NMR spectrum	Confirmed
	HPLC- Mass	Confirmed
Solubility in water	—	—
pH	—	—
Melting point	—	—
Heavy metal	20ppm max	Passed
Arsenic	2ppm max	Passed

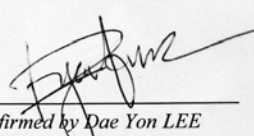
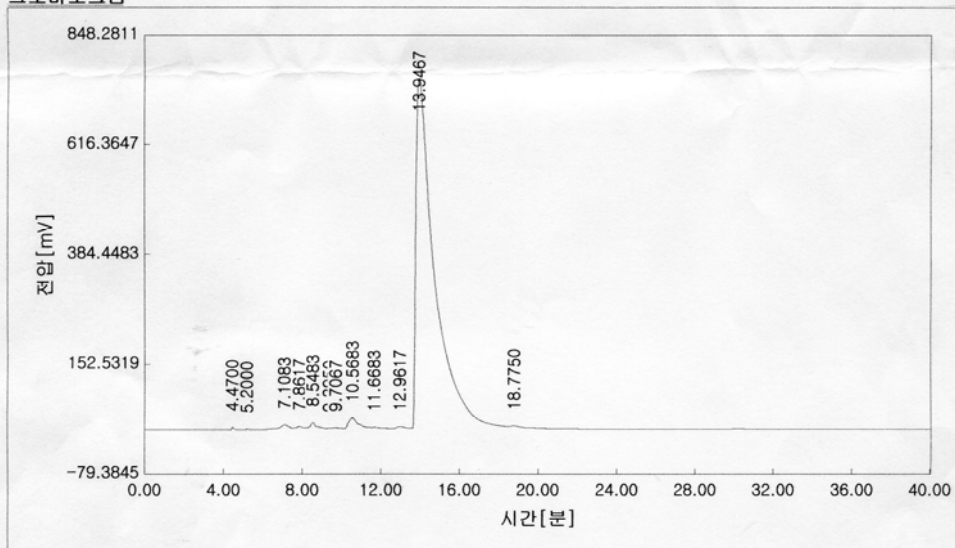

Confirmed by Dae Yon LEE
Senior Research Scientist
Leadgenex, Inc. Research Center

Figure S8. Certificate of analysis for 43 provided by Leadgenex Inc.

부연 설명

Column : Agilent Prep-C18 / 21.2 x 150mm, 5um
Solvent : H₂O/Acetonitrile=30/70
Flow Rate : 5mL/min

크로마토그램



분석 결과

번호	RT[분]	면적[mV*s]	형태	폭[초]	면적%
1	4.4700	32.0213	BB	18.1000	0.0633
2	5.2000	5.8482	BB	10.7000	0.0116
3	7.1083	183.7031	BB	50.1000	0.3631
4	7.8617	47.1636	BB	28.6000	0.0932
5	8.5483	120.3895	BB	39.1000	0.2379
6	9.3950	2.4834	BB	12.0000	0.0049
7	9.7067	12.3668	BB	21.1000	0.0244
8	10.5683	435.8453	BB	44.1000	0.8614
9	11.6683	13.4259	BB	23.0000	0.0265
10	12.9617	97.3601	BB	46.0000	0.1924
11	13.9467	49599.3871	BB	269.7000	98.0317
12	18.7750	45.2488	BB	39.0000	0.0894
합계		50595.2429			

Figure S9. HPLC data for 43.

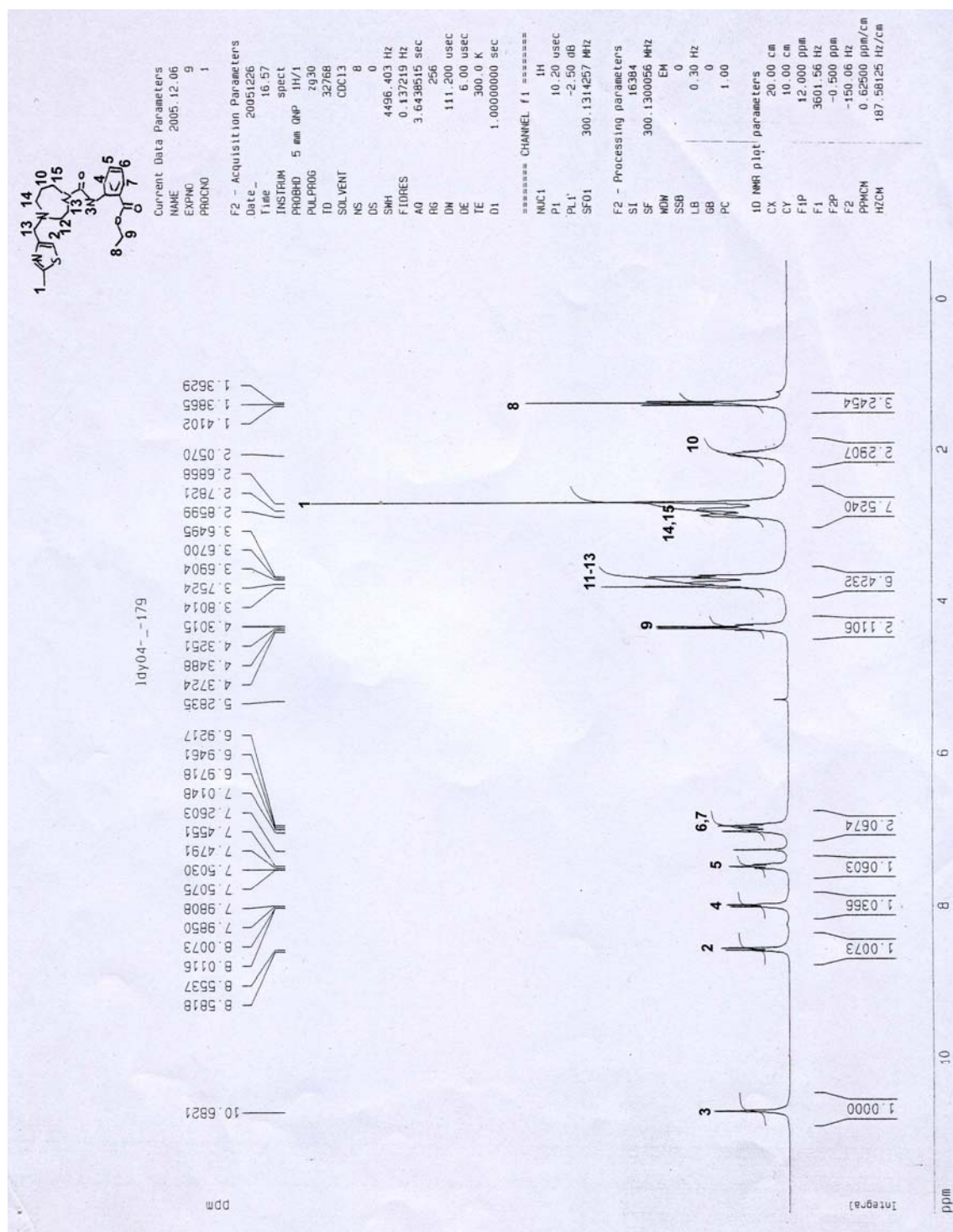


Figure S10. ^1H -NMR spectrum of 43

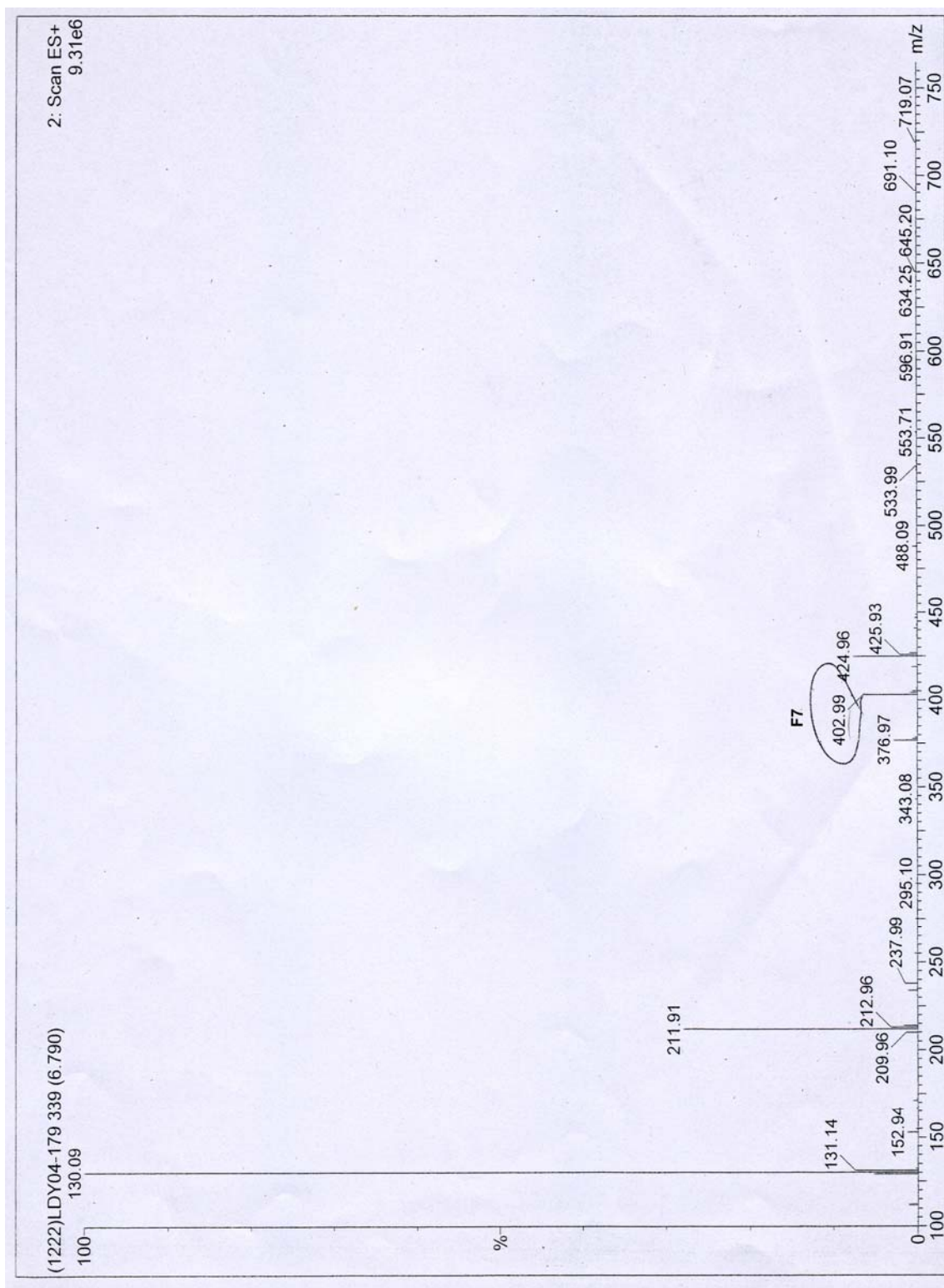


Figure S11. MASS spectrum of 43

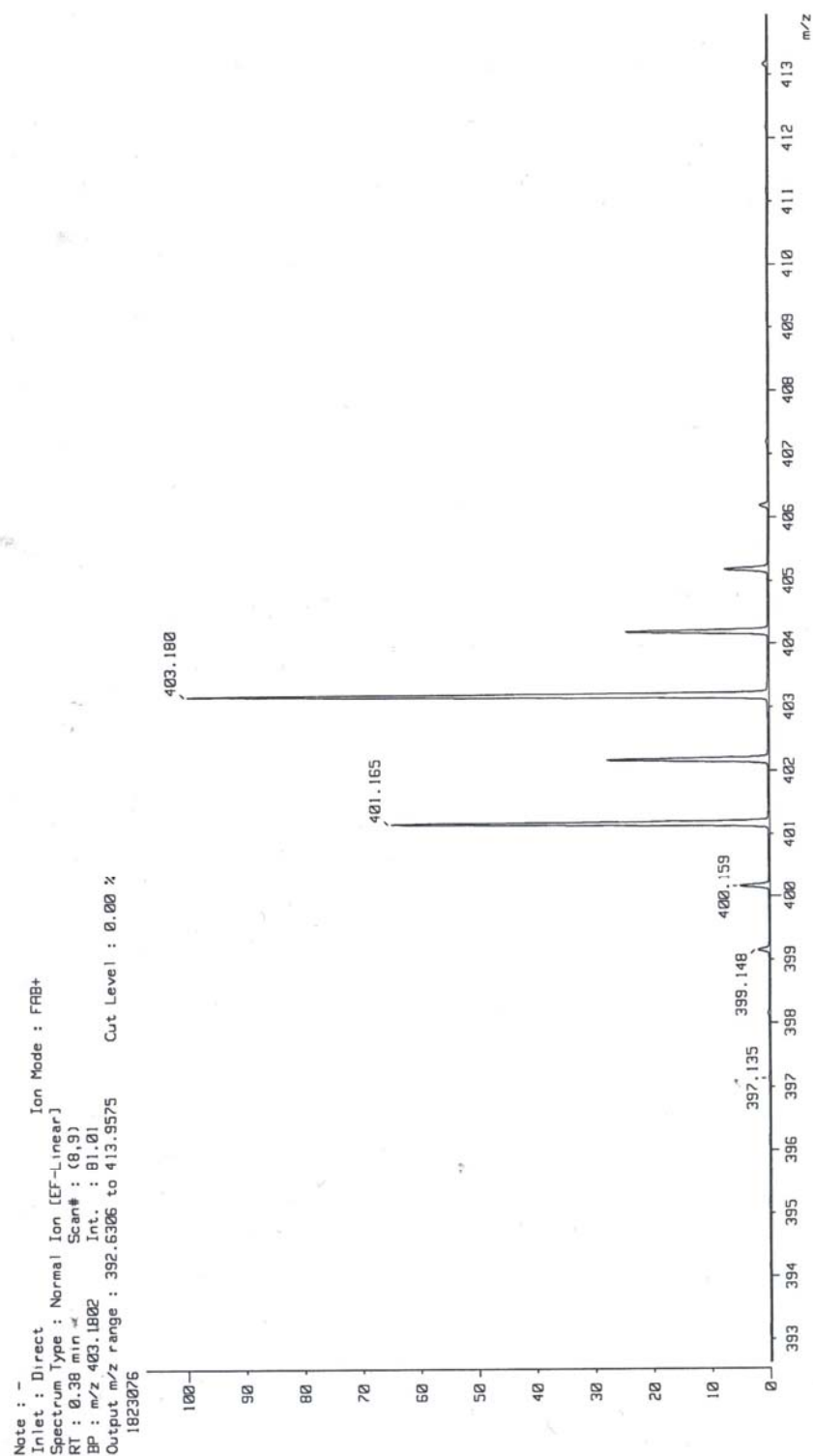


Figure S12. HR_FAB MASS spectrum of 43. The mass of **43** was confirmed. Observed mass corresponds to $[M + H]$. HR FAB-MS for $C_{20}H_{27}O_3N_4S$ m/z calcd 403.1804, found 403.1802.

References

- (1) Nix, J.; Sussman, D.; Wilson, C. *J. Mol. Biol.* **2000**, 296, 1235-1244.
- (2) Nixon, P. L.; Rangan, A.; Kim, Y.-G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P. *J. Mol. Biol.* **2002**, 322, 621-633.